General Guide to Census Data:

If your study area is a city, check the local and state government GIS websites before resigning to downloading data from the Census. Often populated cities will have dedicated GIS departments that will convert recent Census data in a useable spatial format. It is worth checking this to see if you can save time.

If this is not an option, you must deal with the dreaded Census website.

Census data have two components: spatial and attribute. You must download them separately, then join them together.

1. Spatial Component:

Go to https://www.census.gov/cgi-bin/geo/shapefiles/index.php

Census spatial data are called 'TIGER/Line Shapefiles'. TIGER stands for Topologically Integrated Geographic Encoding and Referencing. These Shapefiles will give you spatial boundaries for your study area, but are empty of relevant data except for a linking key field.

Select 2020 (or year of interest) as the *Year*

Layer Types is where you define what geography you want to use:

- Blocks are smallest unit, but do not have demographic data for privacy reasons
- Block Groups are second smallest and are the most granular for demographic data
- Tracts
- ZCTAs- zip code areas
- Counties
- School districts
- Many other options

Regardless of what geography you choose, you will likely be downloading the Shapefile at a state-wide scale, even if you only want to focus on one city. This is fine- we will narrow down the study area after our data are joined together.

Click on the *Submit* button.

On the next page, choose your state of interest from the *Select a State* dropdown, then click *Download*. A zipped Shapefile of the census spatial data will download to your *Downloads* folder. It is a good idea to copy it over into whatever working folder you are using.

Unzip the downloaded zip file to be able to use it.

In QGIS, add the downloaded Shapefile .shp to the map. **Project** it from GCS to whatever projection is best for your study area. Remove the original spatial data layer, leaving just the projected layer to work with.

2. Attribute Component

Go to https://data.census.gov

Under the search bar, click on *Advanced Search*

Under **Find a Filter**, click on *Geography*

Under **Geography**, click on your geographical unit of analysis (Block Group, Tract, etc.)

Under **Select State**, click on the state that your study area is within.

Under **Select County**, either select the county that your study area is within to limit your search, or check "All <Geographical Units> within <Your State>"

Click on the *Search* button.

Under **Tables**, scroll to see different options.

Census data is collected using different surveys. The Decennial Census happens once every ten years and surveys the entire country. The American Community Survey (ACS) usually happens every year, but is an estimation based on sample points. Sometimes the error from the estimation is large. It is usually safest to use 5-year ACS data to reduce this error, unless you are specifically looking at time period.

For demographic data, the Census usually releases separate tables for each topic (e.g.; one table for Race, one table for Age and Sex, one table for Median Household Income, etc.). You will need to download each table separately, edit the tables, then join them all to the census spatial data.

Click on the table with your topic of interest (try to pick the table with the broadest title for that topic). Some of these tables will also report a Margin of Error to give you an idea of how accurate the data are. If you are only concerned with population totals, deselect the Margin of Error button to be more efficient. If this is not an option, continue to the next step.

Click the *DOWNLOAD TABLE* button. If there is no button for this, click on the 'CSV' icon.

Choose to download the most recent dataset, or the year that matches your analysis.

Copy the downloaded zip file to your working folder.

Unzip the downloaded attribute dataset zip file. You will see three files. We want to use the largest file, usually with 'Data' at the end of the name.
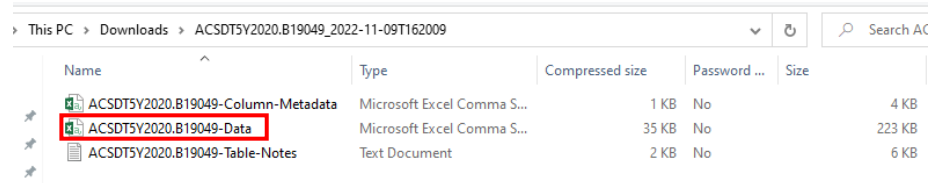
Open this table in Excel.

The format of your table will look different

depending on which table you download. The typical format for Census tables has GEOIDs or identifying codes for each geography feature in the leftmost column, followed by columns with estimated totals for different groups. The top field names of your table use census acronyms and codes, but don't mean anything to an average user. The second row has the meaningful name of each field.

	A	B	C	D	E	F
1	GEO_ID	NAME	B19013_001E	B19013_001EA	B19013_001M	B19013_001MA
2	Geography	Geographic Area Name	Estimate!!Median household income i	Annotation of Estimate	Margin of Error!!Median	Annotation of Margin of Err
3	1500000US330019651001	Block Group 1, Census Tract 9651, I	90217	null	12965	null
4	1500000US330019651002	Block Group 2, Census Tract 9651, I	67992	null	11631	null
5	1500000US330019651003	Block Group 3, Census Tract 9651, I	70938	null	23421	null
6	1500000US330019652001	Block Group 1, Census Tract 9652, I	92266	null	21281	null
7	1500000US330019652002	Block Group 2, Census Tract 9652, I	44643	null	15981	null
8	1500000US330019652003	Block Group 3, Census Tract 9652, I	33610	null	7202	null
9	1500000US330019653001	Block Group 1, Census Tract 9653, I	77885	null	30224	null
10	1500000US330019653002	Block Group 2, Census Tract 9653, I	42917	null	31329	null
11	1500000US330019653003	Block Group 3, Census Tract 9653, I	-	-	**	**
12	1500000US330019654001	Block Group 1, Census Tract 9654, I	66283	null	5843	null
13	1500000US330019654002	Block Group 2, Census Tract 9654, I	87411	null	40551	null
14	1500000US330019655981	Block Group 1, Census Tract 9655.9	55609	null	21576	null
15	1500000US330019655982	Block Group 2, Census Tract 9655.9	82813	null	43028	null
16	1500000US330019655983	Block Group 3, Census Tract 9655.9	41888	null	11611	null
17	1500000US330019656001	Block Group 1, Census Tract 9656, I	81114	null	10123	null

Widen the fields or click on the title cell to see the full name of the fields.

If there is not a second row of meaningful names, there will also be a file in the original downloaded zip file with 'Metadata' at the end. This file lists the field name codes and describes what information they contain. You would then need to use the Metadata file for reference when deciding what fields to use and what to rename them.

If your table does not look like the screenshot above, please go to the Appendix on page 6 for different instructions

Create a new sheet in Excel to receive the demographic data you are interested in. This will be the spreadsheet we eventually bring into QGIS.

Copy over the **GEO_ID/Geography** field. The easiest way to do this is to click on the column heading letter in the gray area above the cells so that everything is selected, then ctrl c and ctrl v into the new worksheet. Don't worry about heading names for now- we will fix that later.

Copy over other fields relevant to your study. These will usually start with !!Total!!, Estimate!!Total or Estimate!!<Your Factor of Interest>. You should probably copy over whatever field holds the total summation for that topic, along with more specific total numbers for sub-topics (ex: For a table showing Age & Sex, copy over Estimate!!Total: column for total population and Estimate!!Total:!!Female:!!18 and 19 years column for total females between 18-19 years old).

Continue until all fields of interest are copied over.

Save the new Excel sheet into a csv file, choosing the format *CSV (MS DOS)*. Give it a name that you will remember.

In the CSV, we will now edit the **GEO_ID** to match our IDs from the Census spatial dataset. Select the entire GEO_ID column, then click on the **Find & Select** tool from the Home tab -> Editing section. Click on **Find** from the dropdown menu.

In the "Find what:" field, type **1500000US**. Click on the gray **Options** box, then check the box next to **Match Case**. Click the **Find All** button. You will see a list pop up at the bottom with all cells containing that string of numbers.

Go to the **Replace** tab and click the **Replace All** button. Here, you are using the Replace tool to replace "1500000US" with nothing, i.e., removing the front part from the ID number. Click **OK**. Select the entire column, then change the format of the ID column to **Number** (In the Home tab -> Number section -> Drop down menu at the top).

If there are decimal digits after the number, click on the icon with three zeros and a blue back arrow in the Number section of the upper ribbon (image on right). This will remove one decimal place at a time. Use it until you no longer have any decimals.
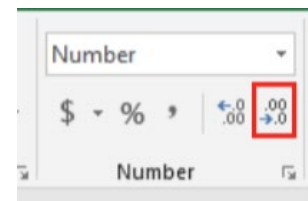
After this, all the IDs should be free from '1500000US'.

Change the field names in the **second row** (the meaningful ones) to ones that are acceptable to QGIS. Remember that you CANNOT use spaces or underscores. Try to keep the field names short, since the software will not accept long names. (e.g.; Estimate!!Total:!!Female:!!18 and 19 years could be renamed something like F18_19).

If it is helpful, open a new Excel sheet and write down the original field names for each column, the updated acronym you gave it, and what data it contains. This will be handy if you have to go back into your data later and can't remember what your renamed field names mean.

Continue changing all the field names until the entire second row has been edited.

Delete the first row (the non-meaningful names).

Save the new sheet and close it in Excel.

If you have downloaded multiple tables from the Census and want to combine them, it is probably a better idea to combine them in QGIS as separate tables than to combine them within Excel. The data records may be in different orders within the Excel table, so you could be copying items to a GEO_ID that actually belong to another GEO_ID. QGIS will address this issue during our joining process, so we don't have to worry.

Go through the same data table editing and renaming steps for each topic table. When all of them have been converted to CSVs with trimmed ID numbers and edited field names, close Excel and go back to QGIS.

3. Joining Census Spatial and Attribute Data

In order to join the spatial and attribute data, we must have a common linking key. However, our GEOID field within the tiger line shapefile is a Text data type, while our CSV file is numerical. We need to copy over the GEOID text values into a new Integer type field, and use that for the join to work.

In QGIS, open the attribute table of your spatial Census dataset. Start a new editing session by clicking the pencil icon. Open the field calculator icon from the top (abacus), and make sure the box is checked next to 'Create a new field'. Give the output field name something like **ID** and set the *type* to be **Whole number (Integer 64 bit).** Make sure the length is greater than the default of 10- maybe around 15 or 20 depending on how long the values in your GEOID field are (this length is only possible if the field type is integer 64 bit). In the expression box, type in the name of the GEOID field surrounded by quotes ("GEOID"), then hit OK.

You should see a new field at the far right of the table with GEOIDs copied over. It may have commas within the numbers, which is fine. Make sure to save your edits by clicking on the pencil icon before moving on.

Add the CSV to your QGIS project using the Data Source Manager's Delimited Text tab. Make sure the file format is CSV, and the geometry definition is set to 'no geometry'. Then join the CSV to the spatial Census Shapefile with the Join Attributes by Field Value tool. Make sure the first input layer is the spatial Census Shapefile, not the table (otherwise we will join to the table). Also make sure to save the output file permanently on your drive.

When the tool finishes, there should be a message telling the number of features that were matched and not matched. There should be more than 0 features matched. Open the attribute table for the new output file to double-check that it went through. If it worked, you should have new fields with new numbers. If it did not, you will likely see the new fields entirely full of null values.

Important note: you will see null values even when your join worked if your spatial Census Shapefile and attribute tables are at different scales (e.g.; your Shapefile has all block groups within a state and your attribute table has demographic values for only one country). Click on the new ID field multiple times to sort the values by ascending and descending, or quickly scroll through all its records. You should see some ID values among all the null values.

If your join failed and there are no values:

- This could be due to a data type mismatch.
  - Go back into CSV in Excel and make sure the formatting for your ID column is set to 'Number'. Check the fields view for your attribute table and make sure the data type for the new ID field is set to 'Double'.
- The columns could have been copied incorrectly. Compare the ID values from your CSV to those in the originally downloaded Census table. If they do not match, copy over the GEOID/Geography field again, change its formatting, and try to join again.
- QGIS might be running into issues dealing with the scale. Sometimes the join won't work in QGIS, but will work in ArcGIS Pro. Then you can export the joined file and continue working with it in QGIS. Reach out to Aletha if you don't know how to use ArcGIS Pro and think this would help.

Export the joined Census dataset to make the join permanent. This is a necessary step if you will be doing any analysis with the Shapefile, since running tools on tenuously joined tables can cause errors. Name the output dataset something you will remember.

If you do not want the entire state for your analysis, you can then either clip the dataset (use Clip tool from the toolbox with the joined layer as your input layer and a study area boundary dataset as your

overlay layer), or select certain counties or regions that you can export into a separate dataset (hand select or select by value).

<u>Also important to note:</u> The Census is fundamentally flawed in how it outlines certain questions, particularly erasing Hispanic/Latino people, biracial people, and LGBTQ+ people. It has also historically undercounted populations that are hard to reach through surveys, which are typically lower income communities of color, undocumented immigrants, or members of tribal nations. Although it is the best representation of US demographics available and is used as a fundamental data set, it is important to recognize its limitations.

Appendix

Some problems you might run into:

1. What if my Census data table does not look like the screenshot on page 3?

Depending on what topic and geography you are using, your table may look different. Sometimes it might look like the following:

Label (Grc	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou	Block Grou
Median hc	27,342	75,129	26,250	47,167	84,667	46,000	62,292	107,778	92,538	51,250	59,038	70,302	73,958	59,674	73,507	75,969	37,232	80,250	57,218	77,083	42,917	60,469

This may be because you are zoomed into too a narrow range of geography (only certain counties within a state, etc.) The first thing to try is changing your geography filter to a broader scale (such as all block groups in the entire state). The table format should then change to be more useable. This may slightly slow down processing speeds in ArcGIS, but will save us time in editing the tables and won't cause problems because our Shapefile is already at a state-wide scale.

If changing the filter to a broader geographic scale does not make your table look like the first screenshot, you will need to heavily edit the tables (see following).

2. I don't have a column for GEOIDs

This is troublesome, because we need GEOIDs to then link our attribute data to the Shapefile. But we can work around it. You will likely have field headings similar to the following with number codes corresponding to your geography.

Block Group 1, Census Tract 9601.01, Grafton County, New Hampshire!!Estimate	Block Group 2, Census Tract 9601.01, Grafton County, New Hampshire!!Estimate
27,342	75,129

GEOID codes for block groups are made up of four parts:

State code + County code + Tract code + Block group code

We can piece the GEOIDs together using this information.

Use this page to determine your state and county codes: https://www.census.gov/library/reference/code-lists/ansi.html#county

Scroll until you find the section similar to the image on the right. Then select the state for your study area.

County and County Equivalents
2010 FIPS Codes for Counties and County Equivalent Entities

Select a State	∨

Field Name	Field Description	Example
STATE	State Postal Code	FL
STATEFP	State FIPS Code	12
COUNTYFP	County FIPS Code	011
COUNTYNAME	County Name and Legal/Statistical Area Description	Broward County
CLASSFP	FIPS Class Code	H1

FIPS Class Codes

You will then get a list like the image below. The second column from the left in red corresponds to the state code, while the third column from the left in blue corresponds to the county code.


```
NH,33,001,Belknap County,H1
NH,33,003,Carroll County,H1
NH,33,005,Cheshire County,H1
NH,33,007,Coos County,H1
NH,33,009,Grafton County,H1
NH,33,011,Hillsborough County,H1
NH,33,013,Merrimack County,H1
NH,33,015,Rockingham County,H1
NH,33,017,Strafford County,H1
NH,33,019,Sullivan County,H1
```

Create a new Excel sheet. Make new columns and name them 'State', 'County', 'Tract', and 'Block_Group'.

In the first column 'Block_Group', type in the Block Group code numbers from your field headings. (e.g; from the screenshot of the table above, we would type in '1', '2', etc.) Use the 'Tract' column to type in the corresponding Census tract codes without the decimal points. (e.g.; from our example above, we would type in '960101' for block group 1 and 2).

Next in the 'County' column, type in the County code that corresponds to the county of your block group. (e.g.; in our example, Grafton County corresponds to the code 009). Finally, type the State code into your 'State' column (e.g.; New Hampshire uses the code 33). You can click on a cell, hover the cursor over the bottom right corner, and click and drag to apply the same code to multiple cells at once.

Now create a new field called 'GEOID'. Use the Concat function (improved version of Concatenate) to string these values together in a new field. In the first cell under the GEOID column, type in 'CONCAT(A2:D2)' where A2 and D2 correspond to the first and last cell numbers that we are stringing together for the first row. Press enter, and you should see a GEOID value.

If you do not see a GEOID value and instead only see the CONCAT formula, highlight all the values in the entire column, then change the cell formatting type to 'Number' in the dropdown in the Number section of the upper ribbon. Then try the formula again.
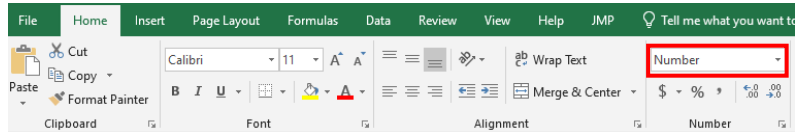
Click and drag the cell with the concatenated GEOID value to the bottom row of the table to apply the formula to all the cells in the column.

Now create new field(s) to hold your data values (values underneath each field heading from the original table). Make sure you are correctly matching the value with the block group code. When you are finished, your table should look something like this:

State	County	Tract	Block_Group	GEOID	Median_HH_Income
33	9	960101	1	330099601011	27,342
33	9	960101	2	330099601012	75,129
33	9	960102	1	330099601021	26,250
33	9	960102	2	330099601022	47,167
33	9	960102	3	330099601023	84,667
33	9	960200	1	330099602001	46,000
33	9	960200	2	330099602002	62,292
33	9	960200	3	330099602003	107,778
33	9	960200	4	330099602004	92,538
33	9	960300	1	330099603001	51,250

Double-check to make sure that your GEOID column has the data type 'Number' so that it joins correctly to your Shapefile.

Now your attribute table should be good to join, so go back to Page 4 Part 3 to finish the process.